ISSN 1561-5359. Штучний інтелект, 2017, № 1

УДК 004.8

В.О. Кузнєцов
Інститут кібернетики імені В.М. Глушкова НАН України, Україна
пр. Академіка Глушкова, 40, м. Київ, 03680

РОЗРОБКА ОСНОВНИХ ЕЛЕМЕНТІВ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ АНАЛІЗУ МІМІЧНИХ ПРОЯВІВ ДЛЯ СИСТЕМ ІНТЕРАКТИВНОГО ВИВЧЕННЯ ЖЕСТОВОЇ МОВИ

V.A. Kuznetsov
V.M. Hlushkov Institute of Cybernetics of NAS of Ukraine, Ukraine
40, Academician Hlushkov av., Kyiv, 03680

DEVELOPMENT OF MAIN COMPONENTS OF INFORMATION TECHNOLOGY FOR FACIAL EXPRESSIONS ANALYSIS IN INTERACTIVE SYSTEMS FOR STUDYING SIGN LANGUAGE

У роботі розглянуто основні елементи інформаційної технології аналізу мімічних проявів для системи інтерактивного вивчення жестової мови. Описано основні компоненти інформаційної технології, її експериментальна реалізація. Проведено аналіз впливу кількості ознак, обсягу навчальної вибірки і типу класифікатора на кількість помилок 1 і 2 роду. Удосконалено існуючі алгоритми ідентифікації мімічних проявів шляхом вибору оптимального набору конструктів мультикласифікаторів.

Ключові слова: інформаційна технологія, алгоритми ідентифікації, міміка, жестова мова.

This article describes the elements of information technology for analysis of facial expressions to apply in interactive study of sign language. The main elements of information technology, its structure and experimental implementation has been discussed. Analysis was carried out in order to identify how the classifier error rate depends on type of classifier, number of feature and teach set. Optimal constructs of classifiers were proposed, giving drastic improvement of existing algorithms.

Key words: information technology, algorithm for identification, facial expressions, sign language.

Вступ

Жестова мова – інструмент для передачі комунікаційної інформації в спілкуванні людей із вадами слуху. Чимала кількість людей має вроджені проблеми зі слухом, що ускладнюють або унеможливлюють комунікацію розмовною мовою. В Україні існує державна програма, що визначає основні напрямки із покращення умов людей із вадами слуху, зокрема, планується вивчення жестової мови серед широких верств населення, які найчастіше контактують з людьми із вадами слуху.

Один із кроків, спрямований на вирішення цієї задачі – створення навчальних програм із залученням сучасних інформаційних технологій.

Найбільш повно до вирішення цієї задачі наблизилася навчальна система [4] із використанням просторових тривимірних моделей людини, що інтегрувала засоби і інтерфейси інтерактивного спілкування користувача із комп’ютером, відтворення і розпізнавання вербальної, графічної, текстової і невербальної інформації.

Незважаючи на повноту засобів, що реалізують інтерактивне середовище навчання жестової мови, необхідний розвиток інших засобів відтворення і розпізнавання жестового мовлення, зокрема, мімічних проявів, які в жестовій мові позначають інтонацію, наголос, логічний і емоційний відтінок речення. Розпізнавання таких елементів жестової мови є важливим, оскільки ці елементи в розмовній мові передаються голосом, натомість в жестовій мові ці аспекти мовлення передаються частково за допомогою видозміни жесту (амплітуда і плавність рухів), що для навчання не є зовсім прийнятним, оскільки жести з високою швидкістю

погано сприймаються. Для полегшення сприйняття, жест супроводжується мімікою, яка, як правило, використовується в калькуючій жестовій мові (що повністю повторює граматику розмовної мови до послідовності всіх граматичних конструкцій включно), або у носіїв жестової мови із дуже високим рівнем знання жестової мови (є аналогом виразності в розмовній мові).

У попередніх дослідженнях [5-7] нами було напрацьовано алгоритми і методи, що можуть застосовуватись в інформаційній технології аналізу (ідентифікації) мімічних проявів на обличчі людини; зокрема, було розглянуто послідовність отримання відеофрагментів мімічних проявів із залученням оптичних давачів (маркерів), закріплених на обличчі людини [5], питання застосування алгоритмів комп'ютерного зору для отримання числових даних, що описують зміну стану мімічних проявів у часі [6], засоби аналізу числових даних, зокрема, одношарового перцептрону та методів зменшення розмірності даних [7].

Результати дослідження [7] показали, що існує необхідність покращення ефективності роботи алгоритмів для задачі ідентифікації зміни мімічних проявів у часі; існуюча реалізація демонструвала невисокий рівень роздільності різних класів мімічних проявів. З метою удосконалення існуючих результатів пропонується наступна постановка задачі:

- запропонувати набір характеристик, які можуть використовуватися для ідентифікації мімічних проявів на підставі даних, отриманих із відеопотоку;
- проаналізувати існуючі засоби і методи розпізнавання мімічних проявів на наборі характеристик, що застосовуються при аналізі мімічних проявів та в інших задачах, що зводяться до ідентифікації зміни характеристик об'єкта в часі;
- розробити експериментальну інформаційну технологію ідентифікації мімічних проявів та її програмну реалізацію;
- провести експериментальні випробування існуючої інформаційної технології з метою встановлення найкращих налаштувань алгоритмів, їх комбінацій та умов, за яких можливо покращення результатів, отриманих раніше.

1. Методи ідентифікації зміни характеристик у часі

Огляд методів, що застосовувалися у попередніх дослідженнях. У роботі [7] було встановлено, що для задачі ідентифікації миттєвого стану обличчя достатньо ідентифікувати стан спокою і відрахувати величину відхилення всіх складових мімічного прояву або мімічного виразу від стану спокою, що не вимагає складних алгоритмів. Цілком достатньо алгоритму ідентифікації на базі перцептрона чи методу опорних векторів, які дають цілком прийнятну якість розпізнавання – порядку 80-90% у найгіршому випадку – малій і неінформативній навчальній вибірці, малому числу ознак і до 99,5% при достатньому розмірі навчальної вибірки. Натомість, ідентифікація тривалого прояву від стану спокою до стану насичення, навіть для двох класів мімічних проявів, дуже ускладнена і потребує ретельного дослідження. Аналогічне спостерігається і при застосуванні методів зменшення розмірності даних до даних, що описують миттєвий стан мімічного прояву. Це вказує на подібність джерел спотворень, що призводять до помилок ідентифікації.

Відмінність аналізу миттєвого стану від послідовностей станів. При аналізі змін характеристик об'єкта в часі, протягом тривалості мімічної активності, прояви мають різну інтенсивність початку і закінчення, різну амплітуду, різну форму сигналу і спотворення сигналу, різну початкову фазу, форму фронту і спаду сигналу та різну тривалість.

Методи дослідження часових послідовностей. Для розв'язання цієї задачі (аналіз послідовностей станів) найбільш прийнятним і доцільним є знаходження подібних задач і алгоритмів розв'язання таких задач, що оперують ідентифікацією

послідовностей станів (для яких алгоритми ідентифікації на подібних даних дають прийнятну точність).

Найбільш близькими є декілька видів задач:

- аналіз фонограм (спектрограм);
- аналіз нейроімпульсів мозку, аналіз кардіограм, міограм;
- аналіз процесів в окремих нервових клітинах (т.з. neuralspikes).

Для розв'язання кожної із задач використовуються свої методи (що найбільш уживані дослідниками).

У сучасних дослідженнях, що пов'язані з аналізом фонограм [8], застосовуються методи глибокого навчання (як із вчителем, так і без вчителя) – згорточні нейронні мережі, знешумлюючі багатошарові автоенкодери, згорточні автоенкодери, глибока мережа переконань тощо. Слід також відзначити метод динамічної згортки [9], що раніше застосовувався для порівняння кореляції двох близьких за формою звукових сигналів різної тривалості. Для дослідження електрокардіограм та іонних каналів використовується перетворення Карунена-Лоева [10] й інші інтегральні перетворення.

Вибір методів аналізу часової послідовності. Пропонується створити алгоритмічну реалізацію для наступних методів:

- *методу динамічної згортки сигналу* щодо чистого вихідного сигналу для зразків мімічних проявів різної розмірності і побудову класифікатора, що використовує міру кореляції як між сигналами відповідних давачів різних сигналів, так і від давачів між собою всередині одного об'єкта, на основі SVM, ШНМ та одного із методів глибокого навчання. Це, зокрема, дозволить перевірити гіпотезу про породження мімічних сигналів одним нервовим імпульсом (і відповідну їх синхронність) через оцінку кореляції сигналів різних компонентів мімічного прояву;
- *методів аналізу спектру* до сигналу, отриманого шляхом приведення вихідного сигналу до фонограмо-подібного вхідного сигналу із визначенням параметрів спектрального аналізу, побудови спектральних характеристик для зразків сигналів та побудову класифікатора на базі згорточної нейромережі. Це дозволить перевірити можливість згортки сигналу вікнами фіксованої розмірності, порівняно із динамічною згорткою.

Крім того, будуть удосконалені існуючі реалізації, що включають у себе інтегральні перетворення, зокрема, перетворення Карунена-Лоева для аналізу сигналів фіксованої довжини і нормованих по амплітуді та дискретне косинусне перетворення [11], а також реалізація алгоритмів зменшення розмірності даних і класифікації для аналізу миттєвого стану мімічних проявів.

2. Інформаційна технологія аналізу мімічних проявів

Інформаційна технологія описує процес обробки даних, що містять мімічні прояви. З метою отримання програмної процесу реалізації, що реалізує цей процес, необхідно визначити усі складові такого процесу – об'єкти, властивості і стани об'єктів, взаємозв'язки і залежності на різних етапах обробки даних.

Вхідні дані. Вхідними об'єктами виступають відеозразки мімічних проявів (відеопотік), які містять послідовність станів мімічного прояву (мімічного жесту або мімічної морфеми) упродовж інтервалу м'язової активності (скорочення-розслаблення одного упродовж інтервалу м'язової активності (скорочення-розслаблення одного мімічного м'яза або групи м'язів).

Атрибутами такого сигналу виступають основні інформаційні характеристики – частота дискретизації сигналу, кількість інформаційних відліків сигналу, порядковий номер (ідентифікатор) сигналу, ідентифікатор мовця тощо.

Між елементами вхідних даних виникають зв'язки ієрархії: кожен відлік сигналу є частиною множини відліків із певним порядковим номером, що визначає черговість появи конкретного відліку у потоці відео.

Обробка вхідних даних. Процес обробки вхідних даних оперує потоком кадрів відео: за допомогою алгоритмів ідентифікації і відслідковування об'єктів на відеопослідовностях отримуються координати ключових точок на кожному кадрі і формується потік даних координат ключових точок на відео.

На даному етапі необхідна наявність обмежень до вхідних даних, що описуються такими термінами: кут орієнтації обличчя відносно камери, швидкість показу мімічних проявів, тип відеокомпресії, освітленість сцени, а також обмежень до алгоритму відслідковування об'єктів, що залежать від типу і налаштувань алгоритмів.

Отримання траєкторій руху ключових точок. Потік даних на відеозображенні можна розглядати у вигляді дискретизованої траєкторії руху точок у деякому просторі, утвореному шляхом геометричних перетворень (у т.ч. проективних) простору мімічних проявів зі зміною масштабу, повороту і зсуву базової системи координат.

Оскільки кожному елементу (відліку) траєкторії руху ключових точок ставиться у відповідність кадр відеопослідовності, то можна згрупувати вхідні дані на деяких характеристиках, отриманих шляхом аналізу траєкторій руху ключових точок (наприклад, максимум і мінімум активності, мінімум, середнє значення за проміжок часу та ін.) та кадрів вихідної відеопослідовності.

Розмітка даних. Даний процес описує створення деяких метаданих, що дозволяють згрупувати кадри відеозображення і відповідні їм стани ключових точок на обличчі. Для розмітки даних використовується автоматизована обробка відеофрагментів і відповідних їм даних про траєкторії руху. На виході процесу обробки даних виникає впорядкований список номерів кадрів вихідної відеопослідовності, і відповідних їм позначень (наприклад, кадр, що відповідає стану спокою і кадр, що відповідає деякій мімічній активності), що дозволяє оперувати конкретними миттєвими станами обличчя, які є найбільш інформативними.

Нормалізація даних. Для подальшої обробки впорядкованого списку кадрів необхідно привести усі зразки рухів до одного масштабу. З цією метою, для всіх зразків, які було знято для конкретного актора, дані переводяться у нову метрику, яка буде співставна для усіх акторів, що показували мімічні прояви, і дозволяє врахувати амплітуду рухів кожної мімічної морфеми, яка є індивідуальною для кожного актора.

Налаштування функціонального перетворювача. Даний процес описує ітераційну процедуру визначення зв'язків (функціональних залежностей) між вхідними даними та деякими лінгвістичними змінними, що описують приналежність вхідних даних (відеозразків мімічних проявів) до деякого класу. Оскільки для вхідних даних є можливість встановлення апріорного класу (заздалегідь відомого значення лінгвістичної змінної), можна побудувати деякий функціональний перетворювач, що оперує даними траєкторій руху ключових точок на обличчі.

Найбільш пристосованими для цієї задачі є алгоритми класифікації і алгоритми навчання цих алгоритмів, що реалізують ітераційну процедуру встановлення зв'язків.

Вихідні дані. Вихідними даними виступає потік значень лінгвістичних змінних, що відповідають деяким відлікам вхідної відеопослідовності або усьому інтервалу відліків. Для таких даних можливо встановити деякий атрибут, що описує точність встановлення функціональної залежності (ідентифікації) для деякого конкретного випадку (мімічний вираз або мімічна морфема) шляхом багаторазового виконання ітераційної процедури налаштування функціонального перетворювача

для декількох реалізацій даних, що сформовані із усієї множини відеозразків мімічних проявів.

Реалізація технології. На основі запропонованої структури створена експериментальна реалізація технології у середовищі IntelliJ IDEA [12] мовою Scala [13] із залученням бібліотек Java. Для реалізації операцій збереження даних створено базу даних у системі керування базами даних PostreSQL [14] та використано відповідний драйвер бази даних для Oracle JVM (Java Virtual Machine) [15], що підключається до виконавчих файлів експериментальної програмної реалізації.

Докладніше про результати роботи експериментальної реалізації у розділі «Експериментальні випробування інформаційної технології ідентифікації мімічних проявів».

3. Експериментальні випробування інформаційної технології ідентифікації мімічних проявів

Вхідні дані. Випробування інформаційної технології було проведено на наборі з 300 мімічних проявів, записаних на відео з облич 5 різних акторів. Дані було розділено на набори, які включали в себе (для різних експериментів) від 2 до 14 представників базових мімічних проявів. Крім того, для ряду експериментів, які включали в себе дослідження одиничного відліку, досліджувався не весь часовий інтервал активності мімічного прояву, а лише той, що охоплював мімічний прояв у стані насичення (приблизно 40-60% усього інтервалу активності).

Отримані зразки мімічних проявів, або їх одиничні відліки із запропонованих наборів розбивалися у співвідношенні 20% для перевірки і 80% для налаштування за допомогою вбудованого генератора випадкових чисел. Ці дані під час кожної ітерації алгоритму кожного разу мінялися таким чином, що кожен елемент даних міг бути використаний і як для перевірки, і як для налаштування.

Опис випробувань. Варіанти реалізацій функціональних перетворювачів, запропоновані в попередньому розділі («Експериментальна реалізація інформаційної технології ідентифікації мімічних проявів»), досліджувалися в різних умовах. Усього було проведено 5 різних випробувань експериментальної реалізації.

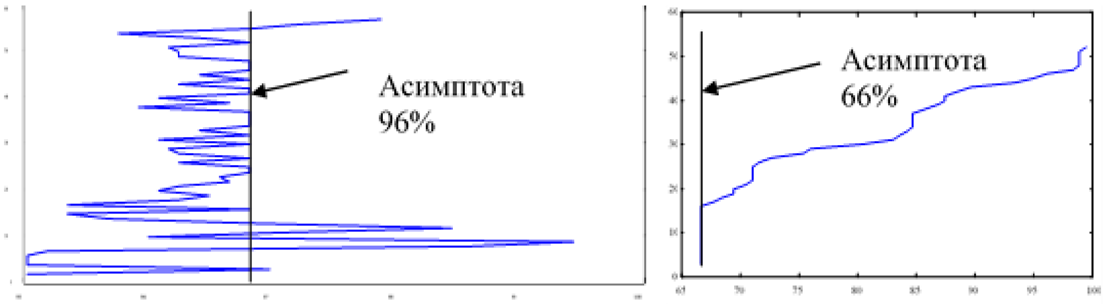
Під час ***першого випробування*** необхідно було виявити, яку інформацію містить зміна мімічних проявів у часі. Для цього було проведено випробування як на окремих відліках даних, так і на цілісних інтервалах активності мімічних проявів у часі. Випробування показало, що зміна мімічних проявів у часі, порівняно із одиничними відліками, містить надлишкову інформацію, яка може бути використана для створення складних гіпотез на підставі вхідних даних, зокрема тих, що враховують інтенсивність змін мімічних проявів. Проте, за різної інтенсивності спостерігалися різні форми сигналу (від гладкої типу «mexican hat» до майже трапецієвидної), що можуть некоректно сприйматися деякими алгоритмами за низької інтенсивності сигналу.

Під час ***другого випробування*** перевірялося, чи застосовне зменшення розмірності даних (як для окремого відліку, так і на цілісних інтервалах активності мімічних проявів у часі). Це випробування було необхідне для встановлення причин великої помилки розпізнавання, отриманої у попередніх дослідженнях [7] із використанням методів зменшення розмірності даних.

Випробування показало, по-перше, що на наявних даних одиничні відліки та головні коефіцієнти розкладення, отримані шляхом зменшення розмірності даних (безпосередньо косинусне перетворення), корелюються між собою, по-друге, що для класифікації даних, отриманих методами зменшення розмірності даних необхідна велика кількість головних коефіцієнтів розкладення (значно більша, ніж 3), що вказувало на те, що власні вектори із низьким рівнем енергії мали суттєвий інформаційний внесок для алгоритмів класифікації (рис. 1.a); при цьому, одиничні

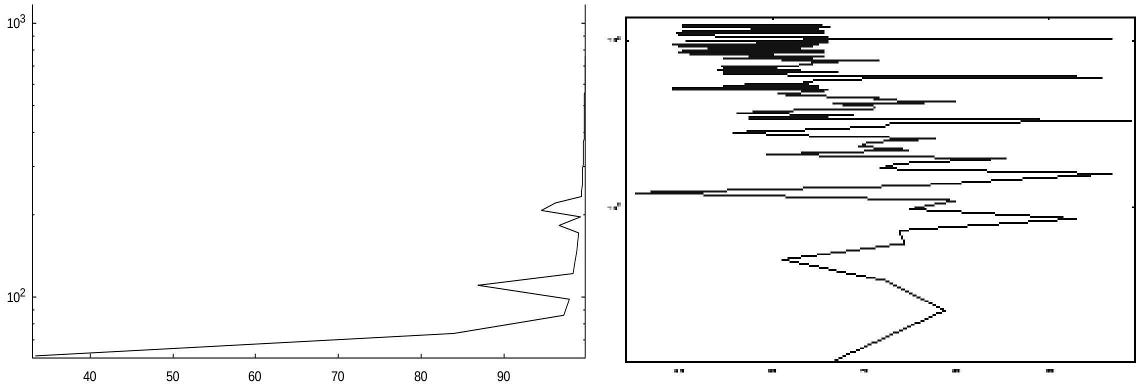
відліки вимагали на один порядок менший обсяг додаткових векторів, аніж для всього часового проміжку мімічної активності – не менше 4 проти 20-40 відповідно.

Це вказує на те, що ефективність алгоритмів класифікації для випадку розкладення цілісного сигналу сильно залежить від кількості коефіцієнтів розкладення методів зменшення розмірності даних, що, в свою чергу, свідчить про велику варіативність даних і сильний вплив фазочастотних властивостей сигналів.



(а) (б)

Рис. 1. Кількість використаних базисних функцій інтегрального перетворення та величина помилки для методу опорних векторів (а) та одношарової нейромережі (б). Вісь ординат – кількість базисних функцій (3-63), вісь абсцис – величина помилки (нижня межа помилки 99,5%).



(а) (б)

Рис. 2. Криві навчання для двох реалізацій класифікаторів: одношарова нейромережа (а) і метод опорних векторів із ядром на основі гіперболічної функції (б). Масштаб логарифмічний. Вісь абсцис – обсяг навчальної вибірки (101-103), вісь ординат – величина помилки (права асимптота ~99,5%, ліва асимптота ~40%).

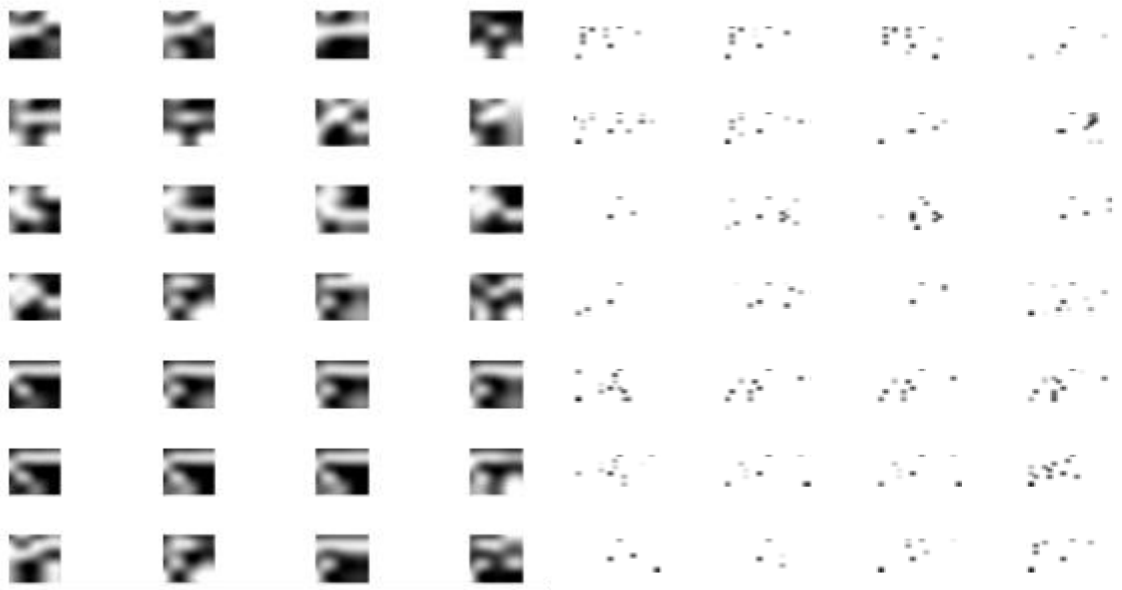
Під час ***третього випробування*** необхідно було виявити мінімальний обсяг вибірки даних, що використовується як для налаштування функціонального перетворювача, так і, зокрема, для перевірки ефективності кожного з конструктів тільки для випадку «одиничний сигнал» (оскільки у попередньому випробуванні вже було показано, що одиничні відліки та головні коефіцієнти дискретного косинусного перетворення корелюються між собою і, відповідно, на вхід класифікатора надходять однорідні дані). Випробування показали, що навіть для обсягу даних порядку 10^3 зустрічаються аномальні викиди (рис.2), що відображають одну з багатьох поганих реалізацій (ітерацій) навчання конкретного алгоритму. Також, виходячи із результатів випробування, можна стверджувати, що величина навчальної вибірки, достатньої для отримання результатів, із задовільною точністю розпізнавання лежить у дуже широких межах: при цьому обсяг вибірки не має такого вирішального значення у величині точності, як самі елементи, які входять у

кожен клас навчальної і тестової вибірок та адаптивність функції до елементів навчальної вибірки. Це вказує на значну дисперсію всередині класу, яка викликає помилки переходу одного елемента класу в інший.

Під час ***четвертого випробування*** необхідно було запропонувати один із шляхів покращення результатів роботи нейронної мережі (яка показала порівняно невелику кількість викидів величини помилки) на чистих вхідних даних, до яких не застосовувалося зменшення розмірності. Одним із засобів є один із методів глибокого навчання, в якому частина даних використовується для попереднього налаштування ваг нейромережі і одночасно як початкова ітерація навчання. У результаті випробування було встановлено, що за допомогою каскадного автоенкодера можна суттєво покращити результати навчання нейромережі із одним прихованим шаром нейронів до порівняння із результатом, отриманим у роботі [7].

Під час ***п'ятого випробування*** необхідно було зменшити аномальні викиди в алгоритмах навчання, що навчаються на даних, отриманих після зменшення розмірності даних. Одним із шляхів покращення роботи алгоритмів є усунення часової, фазової і частотної невідповідності між сигналами, що представляють різні зразки мімічних проявів, без усунення інформації про часові характеристики. Для цього запропоновано порівнювати часові характеристики кожного із мімічних проявів між собою із використанням нормування в часі шляхом динамічної згортки (Dynamic Time Warping). Коефіцієнти незалежної в часі парної кореляції формують власний характеристичний вектор сигналу у вигляді квадратної матриці (рис. 3, а). Після чого, за необхідності, до даної матриці застосовується проріджения із зменшенням розмірності (рис. 3, б). У результаті експериментів було встановлено, що при застосуванні проріджування (одно- та двовимірного), так і без нього, найбільша точність ідентифікації була отримана за допомогою методу опорних векторів.

У результаті проведених експериментів було отримано набір найбільш оптимальних конструктів (комбінацій алгоритмів), які дають порівняно високі показники точності ідентифікації. У таблиці 1 наведено перелік конструктів цих алгоритмів та величину точності ідентифікації на контрольній вибірці для найкращого сценарію (найбільший обсяг вибірки, найкращий набір базисних функцій, найкраща реалізація конкретного алгоритму, тощо).



(а) (б)

Рис. 3. Матриці парної DTW-кореляції для різних зразків мімічних проявів без проріджения (а) та з двовимірним проріджениям за допомогою DCT (б)

Таблиця 1. Точність ідентифікації зразків для різних реалізацій алгоритму

Тип класифікатора	**Розмірність сигналу/кількість ознак при зменшенні розмірності**	
	1x20 (1 відлік часу)	20,30,60,120,…x20 (весь інтервал активності)
SVM (RBF)	99,5%/-	-
ANN MLP	99,5%/-	75%/-
KLT(SVD) + SVM (RBF)	99,5%/1x4	99,5%/1x18
DCT + SVM (RBF)	-	99,5%/3x20
SDAE (400x20x2) + ANN MLP	-	99,5%/-
DBN (400x20x2)	-	75%
ConvolutionalNN 12c-2s-6c-2s	-	75%
DCT+SVM (RBF)	-	99,5%/7x20
DCT+SVM (Linear)	-	99,5%/2x20
DTW + SVM	-	87,5%/-
DTW + ANN/DBN/SDAE/CNN	-	87,5%/-
DTW + DCT + ANN/DBN/SDAE/CNN		87,5%/10x19
DTW + DCT + SVM (RBF)	-	99,5%/7x19
DTW + DCT + SVM (Linear)	-	99,5%/2x19

Висновки

У результаті проведених експериментів покращено технологію ідентифікації мімічних виразів і компонент мімічних виразів, захоплених за допомогою систем захоплення координат керуючих контрольних точок з використання оптичних маркерів, особливо для методів ідентифікації зміни мімічних виразів у часі. Отримано декілька різних структур мультикласифікаторів із комбінацією методів виділення ознак і класифікації мімічних проявів, що дало можливість реалізувати ефективну систему ідентифікації мімічних проявів. Було проведено аналіз впливу кількості ознак, обсягу навчальної вибірки і типу класифікатора на кількість помилок 1 і 2 роду, що дозволило запропонувати декілька нових конструктів мультикласифікаторів.

Література

1. Зайцева Г.Л. Жестовая речь. Дактилология: Учеб. для студ. высш. учеб. заведений / Зайцева Г.Л. – М: Гуманит. изд. центр ВЛАДОС, 2000. – 192 с.
2. Ткачева Т. Жестовый язык глухих как особая знаковая коммуникативная система (на материале русского и французского жестовых языков) / Ткачева Т., Ахмедышева Н. // Материалы Международной научно-практической конференции «Приоритеты и интересы современного общества». – Астрахань: АГУ, 2010. – С. 361–365.

3. Кульбіда С.В. Українська жестова мова як природна знакова система / С. В. Кульбіда // збірник наукових праць «Жестова мова й сучасність»: К: Педагогічна думка, 2009 . – С. 218-239.
4. Крак Ю.В. Розробка основних інструментів комп'ютерної анімації для побудови системи синтезу жестової мови / Ю.В. Крак, М.П. Лісняк, А.С. Тернов // Матеріали Міжнародної наукової конференції «Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту», 27–31 травня 2012. – Херсон: ХНТУ, 2012. – С. 104-105.
5. Крак Ю.В Моделирование эмоциональных проявлений с помощью параметрических моделей головы человека / Ю.В. Крак, А.С. Тернов, В.О. Кузнецов // Науково-теоретичний журнал «Штучний інтелект». – 2013. – № 3. – Донецьк.: ІПШІ, 2013. – С. 154-170.
6. Krak Ju.V. Facial expressions analysis based on computer vision algorithms / Ju.V. Krak, A.S. Ternov, V.A. Kuznyetsov // «International Journal Information Content and Processing». – 2014. – № 4. – Sofia.: ITHEA, 2014. – pp. 326-338.
7. Крак Ю.В Анализ мимических проявлений на лице человека / Ю.В. Крак, А.С. Тернов, В.О. Кузнецов // Науково-теоретичний журнал «Штучний інтелект». – 2015. – № 3-4. – Київ.: ІПШІ, 2015. – С. 37-50.
8. Honglak L. Unsupervised Feature Learning for Audio Classification Using Convolutional Deep Belief Networks // L. Honglak, Y. Largman, P. Pham, Y.Ng. Andrew// Proceedings of conference on Neural Information Processing Systems (NIPS 2009) – Red Hook, NY: Curran Associates, Inc., 2009, - pp. 1096-1104.
9. Al-Naymat G. Sparse DTW: A novel approach to speed up Dynamic Time Warping / G. Al-Naymat, S. Chawla, J. Taheri // The 2009 Australasian Data Mining, vol. 101 – Melbourne, Australia, ACM Digital Library. – pp. 117-127.
10. Chumakov A.G., Kurashov V.N. Karhunen-Loeve Basis Synthesis for Grate Capacity Signal Performance in Optical Processors // SPIE Proc. – 1993. – Vol.21-08. – P. 338- 342.
11. Loeffler C., Ligtenberg A., Moschytz G. Practical Fast 1-D DCT Algorithms with 11 Multiplications // Proc. Int. Conf. on Acoustics, Speech, and Signal Processing 1989 (ICASSP '89). – 1989. – P. 988-991.
12. Електронний ресурс інтегрованого середовища розробки Java/Scala IntelliJ IDEA. Режим доступу: https://www.jetbrains.com/idea/
13. Електронний ресурс мови Scala. Режим доступу: http://www.scala-lang.org
14. Електронний ресурс бази даних PostgreSQL. Режим доступу: https://www.postgresql.org/
15. Електронний ресурс Oracle Java Virtual Machine. Режим доступу: http://www.oracle.com/technetwork/java/javase/overview/index.html

Literatura

1. Zaytseva H.L. Zhestovaya rech'. Daktylolohyya: Ucheb. dlya stud. vsissh. ucheb. zavedenyy / Zaytseva H.L. – M: Humanyt. yzd. tsentr VLADOS, 2000. – 192 s.
2. Tkacheva T. Zhestovy yazik hlukhykh kak osobaya znakovaya kommunykatyvnaya systema (na materyale russkoho y frantsuzskoho zhestovykh yazykov) / Tkacheva T., Akhmedysheva N. // Materyali Mezhdunarodnoy nauchno-praktycheskoy konferentsyy «Pryoryteti y ynteresi sovremennoho obshchestva». – Astrakhan': AHU, 2010. – S. 361–365.
3. Kul'bida S.V. Ukrayins'ka zhestova mova yak pryrodna znakova systema / S.V. Kul'bida // zbirnyk naukovykh prats' «Zhestova mova y suchasnist'»: K: Pedahohichna dumka, 2009. – S. 218-239.
4. Krak Iu.V. Rozrobka osnovnykh instrumentiv komp"yuternoyi animatsiyi dlya pobudovy systemy syntezu zhestovoyi movy / Iu.V. Krak, M.P. Lisnyak, A.S. Ternov // Materialy mizhnarodnoyi naukovoyi konferentsiyi „Intelektual'ni systemy pryynyattya rishen' ta problemy obchyslyuval'noho intelektu", 27–31 travnya 2012. – Kherson: KhNTU, 2012. – S. 104-105.
5. Krak Iu.V Modelyrovanye emotsyonal'nikh proyavlenyy s pomoshch'yu parametrycheskykh modeley holovi cheloveka / Iu.V. Krak, A.S. Ternov, V.O. Kuznetsov // Naukovo-teoretychnyy zhurnal «Shtuchnyy intelekt». – 2013. – № 3. – Donets'k.: IPShI, 2013. – S. 154-170.
6. Krak Iu.V. Facial expressions analysis based on computer vision algorithms / Iu.V. Krak, A.S. Ternov, V.A. Kuznyetsov // International Journal "Information Content and Processing". – 2014. – # 4. – Sofia.: ITHEA, 2014. – pp. 326-338.
7. Krak Iu.V Analyz mymycheskykh proyavlenyy na lytse cheloveka / Iu.V. Krak, A.S. Ternov, V.O. Kuznetsov // Naukovo-teoretychnyy zhurnal «Shtuchnyy intelekt». – 2015. – № 3-4. – Kyiv.: IPShI, 2015. – S. 37-50.
8. Honglak L. Unsupervised Feature Learning for Audio Classification Using Convolutional Deep Belief Networks // L. Honglak, Y. Largman, P. Pham, Y.Ng. Andrew // Proceedings of conference on Neural Information Processing Systems (NIPS 2009) – Red Hook, NY: Curran Associates, Inc., 2009, - pp. 1096-1104.
9. Al-Naymat G. Sparse DTW: A novel approach to speed up Dynamic Time Warping / G. Al-Naymat, S. Chawla, J. Taheri // The 2009 Australasian Data Mining, vol. 101 – Melbourne, Australia, ACM Digital Library. – pp. 117-127.
10. Chumakov A.G., Kurashov V.N. Karhunen-Loeve Basis Synthesis for Grate Capacity Signal Performance in Optical Processors // SPIE Proc. – 1993. – Vol.21-08. – P. 338- 342.

11. Loeffler C., Ligtenberg A., Moschytz G. Practical Fast 1-D DCT Algorithms with 11 Multiplications // Proc. Int. Conf. on Acoustics, Speech, and Signal Processing 1989 (ICASSP '89). – 1989. – P. 988-991.
12. Elektronnyy resurs intehrovanoho seredovyshcha rozrobky Java/Scala IntelliJ IDEA. Rezhym dostupu: https://www.jetbrains.com/idea/
13. Elektronnyy resurs movy Scala. Rezhym dostupu: http://www.scala-lang.org
14. Elektronnyy resurs bazy danykh PostgreSQL. Rezhym dostupu: https://www.postgresql.org/
15. Elektronnyy resurs Oracle Java Virtual Machine. Rezhym dostupu: http://www.oracle.com/technetwork/java/javase/overview/index.html

RESUME

V.O. Kuznetsov

Development of main components of information technology for facial expressions analysis in interactive systems for studying sign language

This article describes the steps for creating the elements of information technology for the analysis of facial expressions, using facial expressions motion-capture systems using optical markers. We analyzed the existing algorithms for the identification of facial expressions, previously developed, as well as the algorithms used for similar tasks (analysis of biological signals).

We have developed structural diagram of the information technology, showing properties, relationships and data flows, but also describes the sequence of processing in the information technology.

On the basis of the structural scheme, we developed experimental implementation of information technology, which included the database based on PostgreSQL and software implementation on Scala.

A series of experiments was conducted on data in order to determine the dependence of the number of classification errors of 1 and 2 kinds on data types (single samples or the entire range of mimic expressions activity), reduction of the data features (using singular value decomposition or not), type of classifier (neural network with or without preliminary choice of weights, support vector machines, etc.), the volume of training samples, the reduction of the “time” coordinate in the data, which (all above) resulted in the plots and score the best indicator of the accuracy of the identification on the basis of which several classificators’ constructs have been proposed.

As a result of experiments, we significantly improved the existing technology of identification of facial expressions and their components, in terms of methods of identification of temporal changes. The set of developed multi-classifiers’ structures, which had noticeable features, was sepated from other structures; this set consist of several most optimal methods of feature extraction and classification of facial expressions, which made it possible to implement an effective system for identifying facial expressions.

Надійшла до редакції 25.04.2017